DOCUMENT RESUME

ED 359 755 FL 020 820

AUTHOR Naslund, Harry
TITLE Identity Relations and Superthemes in Swedish LSP Texts.
PUB DATE Apr 92
NOTE 11p.; In: Nordic Research on Text and Discourse. NORDTEXT Symposium (Espoo, Finland, May 10-13, 1990); see FL 020 811.
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Content Analysis; Economics; Foreign Countries; *Languages for Special Purposes; Semantics; Swedish; *Text Structure; Uncommonly Taught Languages
IDENTIFIERS Sweden

ABSTRACT

This paper addresses the microsemantic and macrothematic analyses of science and popular science texts from the fields of economics, medicine, and technology used in the "LSP [Languages for Specific Purposes] Texts in the 20th Century" project. Results show that the frequency of the relations is different within various parts of the texts. A clear difference between the (pure) identity and the partial identity relations is that the former are the most frequent in theme development, whereas the latter have their highest frequency in the discussion part of several of the text groups. Distribution of possessive relationships indicates that the personal element of the medical texts is concentrated within the introduction and discussion, whereas the various diseases are dealt with in the central theme development. (JP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED359755

Identity relations and superthemes in Swedish LSP texts.

HARRY NÄSLUND

Abstract

An investigation of the distribution of identity relations in superthemes in Swedish economic, medical, and technical LSP texts shows that the frequency of the relations is different within various parts of the texts. As might be expected, identity relations in general are relatively infrequent in introductions. The (pure) identity relations are, on the other hand, comparatively frequent in the theme developments, probably the most homogeneous text parts as regards subject-matter, whereas the partial identity relations have their greatest share in the discussions, particularly in the economic texts, where the discussion parts are relatively theoretical. There is, thus, apparently a connection between the frequency of partial identity relations and the level of theory in the texts. Possessive relationships, which, especially in the medical texts, are often connected with individual human beings, have the highest percentage in the introductions and the discussions, which indicates that human beings are dealt with particularly often in these text parts. In the economic texts, however, the frequency of possessive relationships is highest in the theme developments, which suggests that, in many cases, institutions, e.g. banks, which are here relatively often referred to in the genitive or with a possessive pronoun, are dealt with in this part of the texts.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Tore Ahlbäck

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

FL020820

1. Introduction

The research project 'LSP Texts in the 20th Century' has been in progress at the Department of Scandinavian Languages at the University of Uppsala in Sweden since 1986. The project is headed by Britt-Louise Gunnarsson, who also elaborated its general theoretical framework.

The aim of the project is to investigate LSP texts from semantic and pragmatic points of view and to try to place them in their general and specific contextual frames. The text analysis approach is cognitive, pragmatic, macrothematic, and microsemantic (see Gunnarsson, 1987, 1989).

2. Material

The LSP texts which have been analysed are of two genres, science (written by experts for fellow experts) and popular science (written by experts for lay people), within the fields of economics, medicine, and technology. The texts are from three periods, 1895-1905, 1935-1945, and 1975-1985. For a more detailed description of the material see Gunnarsson, Melander and Näslund (1987).

BEST COPY AVAILABLE

3. Analyses

The cognitive and microsemantic analyses included 90 texts, i.e. five science texts and five popular science texts from each subject and period.

In this paper, I will deal only with certain parts of the microsemantic analysis in combination with some aspects of the macrothematic analysis.

3.1. *The microsemantic analysis*

Our microsemantic analysis is a reference analysis, directed towards the micro level of the text. It is less limited to the surface structure of the text and more cognitive than traditional cohesion analysis, e.g. Enkvist (1974), Halliday and Hasan (1976), and Källgren (1979).

The aim of this analysis was to investigate the semantic relations between words and phrases in the same sentence as well as between words and phrases in different sentences.

We have distinguished four main categories of semantic relation: (1) *identity*, (2) *extension*, (3) *description*, and (4) *causality*, each divided into subcategories. Here, I will confine myself to the identity relations, but a survey of all the relations analysed is presented in Table 1.

Table 1. *Semantic relations employed in the reference analysis.*

Main relations	**Relations**
Identity relations	Identity Partial identity Possessive relationship
Extensional relations	Expansion Diminution Parallelism
Descriptive relations	Main word Description
Causal relations	Cause Consequence Result

3.1.1. *Identity relations*

Among the identity relations we have, as noted in Table 1, distinguished identity, partial identity, and possessive relationship.

Since, as I have mentioned, our reference analysis is a cognitive one, identical reference is fundamental to our identity relations, whereas we consider form to be irrelevant.

By (pure) identity we therefore mean a relationship between coreferential words and phrases, even if the form is different. Thus, identity has been marked when a word is referred to in an identical form, e.g. *sjukdom(en) - sjukdom(en)* ('(the) disease - (the) disease'), and also when it is replaced by a pronoun or another substitute word, e.g. *sjukdom(en) - den* ('(the) disease - it') and vice versa. Also synonymy - the same reference but a quite different form - has been regarded as identity, e.g. *lungsot(en) - (lung)tuberkulos(en)* ('(the) consumption - (the) pulmonary tuberculosis') and vice versa.

Partial identity is the term used when a word has the same reference as part of a preceding compound, e.g. *tuberkulosvaccin(et) - tuberkulos(en)* ('(the) tuberculosis vaccine - (the) tuberculosis') and vice versa, when part of a compound has the same reference as a preceding word, of course also via a synonym or a substitute word. Compounds are very frequent in Swedish.

A possessive relationship is marked when a preceding noun or pronoun in the common case is referred to by a coreferent in the genitive form or by a possessive pronoun, e.g. *sjukdom(en) - sjukdom(en)s* ('(the) illness - of (the) illness'), *den - dess* ('it - its'). Possessive relationship is also used for relations in the opposite direction.

3.1.2. *Content units*

In connection with this reference analysis, each sentence in the text was divided into semantic content units, normally made up of a noun or a verb, sometimes with qualifiers.

After the division into content units, the semantic relations between each content unit and the various preceding units were marked where appropriate. The same content unit can, thus, have relations to various other units, but only one of each main relation, for example, within the identity relations, either (pure) identity, partial identity or possessive relationship.

3.1.3. *Results: the frequency of identity relations in the entire texts*

The scope of the investigation presented in this paper is the distribution of the identity relations in the different superthemes of the texts. The frequency of the microsemantic relations in the texts in their entirety have been accounted for in detail in a special report (Näslund, 1989) and will be presented here very briefly, just to give some background. Only differences between the three fields, economics, medicine, and technology, will be dealt with here.

Table 2 shows the frequency of the three identity relations in the texts in their entirety. All the values are percentages of the number of content units analysed with respect to reference relations.

Table 2. *Identity relations. Percentages of the total number of content units (N=105,867).*

FIELD	ID	PI	PO
Total	26.8	11.5	3.5
Economics	26.4	13.7	4.3
Medicine	26.5	8.7	3.4
Technology	27.6	12.1	2.7

Abbreviations: ID = (pure) identity
PI = partial identity
PO = possessive relationship

As is shown in Table 2, 26.8% of the content units have (pure) identity relations, 11.5% partial identity relations, and 3.5% possessive relationships.

As Table 2 also shows, the percentages accounted for by the (pure) identity relations are almost the same in all the fields, even if the figure is highest in the technical texts (27.6%). These texts seem to be the most internally homogeneous with respect to the contents. This has been shown in other parts of the analyses and presented in reports from the text project (Melander, 1989, and Näslund, 1989). The technical texts are very often about a particular device such as a generator, an elevator, etc. Such homogeneity probably explains the relatively high percentage of (pure) identity relations in technology, since the same concepts then occur many times in the same text.

Table 2 shows in addition that the partial identity relations have their smallest share in the medical texts (8.7%). This is probably connected with the fact that medicine is the oldest of the fields that we have been concerned with and, even during the first period of our investigation, there is, in these texts, an apparently well-established

medical vocabulary, to a great extent consisting of Greek and Latin terms. This has been discussed in the report of the results of the reference analysis (Näslund 1989).

In economics and technology, the levels of partial identity are almost the same (13.7% resp. 12.1%). These two fields are relatively modern, and the development of the fields has given rise to new concepts that have been labelled with new terms, and, in Swedish, many of these terms have been formed by means of compounds.

In technology, these new concepts are very often related to different kinds of innovations, new or better technical devices or parts of these, whereas in economics, the new concepts are mainly due to the development of the field, very much expressed in the form of a raised level of theory. These interrelations have been shown in the report of the reference analysis results (Näslund 1989).

In Table 2, we can also see that the possessive relationships are less frequent in technology (2.7%) than in the other fields, which is not very surprising since these relations are very much connected with human beings, which has also been shown in Näslund (1989), and the technical texts deal very little with persons. By comparison, in the medical texts human beings are more often referred to, and the percentage of possessive relationships is consequently higher (3.4%).

The remarkably high level in the economic texts (4.3%) is very much related to the fairly high frequency of names of countries, institutions, etc. in these texts, since in Swedish such names are often expressed in the genitive or replaced by a possessive pronoun.

3.2. *The macrothematic analysis*

Along with the microsemantic and the cognitive analyses, the 90 texts have also been analysed thematically and pragmatically.

In the thematic part of the analysis, which is directed towards the macro level of the texts, the themes of the longer text parts have been distinguished as four superthemes - introduction, theme development, discussion, and conclusion (see further Gunnarsson, 1989).

3.3. *Identity relations and superthemes in combination*

To establish the distribution of the three subcategories of identity relations in the different parts of the texts, I have investigated how the relations are distributed over the various superthemes. Since some of the texts lack the superthemes of discussion and/or conclusion, these superthemes have been brought together, under the common term of 'discussion', so that almost all the texts have at least one text part with a

concluding supertheme after the theme development. Thus, the three superthemes used in this investigation are *introduction* (I), *theme development* (T), and *discussion* (D).

3.3.1. Results: the distribution of identity relations over superthemes

It is not unreasonable to assume that the percentages of the distribution of the identity relations over the various superthemes will reflect some of the basic variation that exists between the entire texts of the different text groups (Table 2).

To get comparable figures for the levels in both different parts of the same text groups and the same parts of different text groups, the basic variation has been disregarded in Table 3 by giving the percentage of the identity relations in the entire texts of each text group the index of 100. This index thus implies that the frequency of a relation in a text part, classified as a certain supertheme, is exactly the same as the frequency of the same relation in the texts in their entirety, given in Table 2. Figures over 100 accordingly mean that the level of a relation in the text part in question is higher than the level of the same relation in the entire texts, whereas figures under 100 mean that the level is lower. All the percentages of the distribution of the relations over the various superthemes have simply been divided by the percentages accounted for by the same relations in the texts in their entirety and then multiplied by 100. If the percentage of a relation is, for example, 20% in the introduction and the value for the entire texts in the group is 25%, then the index for the introduction will be 20.0/25.0*100=80. All index figures have been rounded off.

Table 3 shows the index figures of the distribution of the three identity relations - (pure) identity, partial identity, and possessive relationship - over the text parts classified as the three different superthemes - introduction, theme development, and discussion.

Table 3. *Identity relations and superthemes.* Index figures show the levels of the three identity relations in the various superthemes compared with the levels of the same relations in the texts in their entirety (=100).

FIELD	IID	TID	DID	IPI	TPI	DPI	IPO	TPO	DPO
Total	92	102	97	85	102	103	100	100	106
Economics	90	103	92	74	101	112	93	105	98
Medicine	90	100	105	108	101	95	121	94	121
Technology	95	101	94	81	103	98	96	100	93

Abbreviations: I = introduction
T = theme development
D = discussion
ID = (pure) identity
PI = partial identity
PO = possessive relationship

As Table 3 shows, the greatest share of the (pure) identity relations is to be found in the theme development (102), whereas the smallest share is in the introduction (92) and the level in the discussion is in-between (97).

It is not very surprising to find the lowest level in the introduction. Here, many concepts are mentioned for the first time, so they cannot have any identity relations to previous concepts. This part of the text might also be more heterogeneous than the other parts (particularly the theme development), as far as text content is concerned, since the introduction refers to circumstances that constitute the background of the theme but do not necessarily have very many identity relations between them.

Neither is it remarkable to find the highest level of identity in the theme development, considering that this text part normally concentrates on the very object of the account, an economic problem, a disease, or a technical appliance, and it is therefore often likely to be the most homogeneous part of the text.

In the discussion, the scope is once again widened with aspects like explanations, recommendations, external consequences, and external actions, and that is probably one important reason why the level of identity is lower here than in the theme development.

If we look at the percentages in the different fields, we find that the relatively homogeneous character of the technical texts in their entirety (Table 2) is to some degree a matter of differences between the fields in the introduction, where the technical texts have a higher level (95) than the economic and the medical ones (90 for both groups). The introduction thus seems to be relatively uniform in the technical texts, which very often deal with only one device at a time.

The relatively high figure in the discussion of the medical texts (105) compared with the other two fields (92 resp. 94) seems to indicate that the medical discussions are more restricted in scope than the economic and technical ones.

As regards partial identity, Table 3 shows that the total figures in the discussion and the theme development are almost the same (103 resp. 102), whereas the level in the introduction is considerably lower (85).

This low level can certainly be explained in the same way as the relatively small share of (pure) identity in the introduction; in the beginning of the text, many new concepts are introduced and they do not necessarily have any identity relations to former concepts.

The somewhat higher level of partial identity in the discussion is probably partly attributable to the interrelation of partial identity and the level of theory, stated above. This level is probably often higher in a text part that deals with such abstract aspects

as, for example, explanation, deduction, consequences, and problem solution, than in the more concrete parts, primarily the theme development.

When looking at the different fields, we can also see that the highest level of partial identity occurs in the discussion of the economic texts (112). In these texts in their entirety, the percentage accounted for by partial identity is the highest (Table 2), and here the interrelation between the partial identity relations and the level of theory, mentioned above, also seems to be the strongest. Since the discussion, as stated above, generally appears to be the most theoretical part of the text, it is therefore not surprising to find the clearly greatest share of partial identity in the discussion parts of the economic texts.

Considering what has already been said about the partial identity relations of the technical texts, it is not unexpected to find the highest figure here in the theme development (103), where some technical appliance or some part of it is described.

In medicine, we find lower levels in the discussion (95) and in the theme development (101) than in the introduction (108). To a great extent, the old vocabulary of Greek and Latin words mentioned before are probably used when referring to the various diseases or cases in the theme development and in the discussion, whereas the presentation of the background, with more theoretical elements, such as previous research, etc., in the introduction, seems to require an enlarged vocabulary, partly made up of new compounds.

When it comes to the possessive relationships, Table 3 shows that the levels are exactly the same in the introduction and in the theme development (100), whereas the value in the discussion is somewhat higher (106). These figures probably indicate that it is very much in the discussion that the phenomena dealt with in the theme development, such as an economic problem, a disease, and maybe also a technical appliance, are related to individual human beings or institutions of various kinds, which, as mentioned above, has been shown to favour possessive relationships.

If we look at the different fields, we can see that this is - as expected - especially true about medicine (121 in the discussion), where human beings also seem to be found in the introductions (121). The primary object of the medical text - the disease - is apparently more related to individual human beings in the introduction, e.g. the background situation, and in the discussion, e.g. the consequences, than in the theme development, where the disease per se is probably the main object of description.

In economics and technology, the possessive relationships are, as stated previously, less connected with human beings than in medicine. The high percentage in the entire economic texts (Table 2), where possessive relationships are often due to the names of various institutions, is very much concentrated to the theme development (105). This probably indicates that, in these texts, institutions such as banks, etc. are more often

dealt with directly in the theme development than indirectly in the introduction and the discussion.

4. Summary and discussion

An investigation of the distribution of the identity relations over superthemes (Table 3) gives a great many results that show differences from the reference patterns of the entire texts (Table 2).

To start with, it is to be expected that the three identity relations have a relatively low frequency in the introduction where particularly many new concepts are brought into the text.

A clear difference between the (pure) identity and the partial identity relations is that the former are the most frequent, in general, in the theme development, whereas the latter have their highest frequency in the discussion part of several of the text groups. More than the other text parts, the theme development deals with the phenomenon in question and is, therefore, relatively homogeneous regarding subject-matter introducing few new concepts, while the discussion may be more abstract and therefore requires a more extensive and advanced vocabulary.

This seems to be especially true about economics, where the development of the field is very much seen in the form of a higher level of theory and a greater percentage accounted for by partial identity in the texts in their entirety. Technology, on the other hand, where the partial identity relations seem to be more connected with various kinds of new inventions, has its greatest share in the theme development, which is completely in agreement with what has been said earlier, since these innovations, which often require new names, are certainly more often described in this part of the texts than in the others.

In medicine, with its old Greek and Latin terms for the various diseases, which constitute the primary subject-matter of the texts and are dealt with in the theme development, the level of partial identity is higher in the introduction where, among other things, more theoretical aspects, such as previous research, are accounted for.

In economics and technology, both relatively new fields, this is not the case. Particularly in the older parts of our corpus there might not always have been so much and such complicated previous research to present.

As to the possessive relationships, their distribution indicates that the personal element of the medical texts is very much concentrated to the introduction and the discussion; human beings are apparently brought into the texts more indirectly, whereas the various diseases, the primary objects of the texts, are dealt with in the central theme development.

In economics, banks and other institutions are often dealt with in the theme development and thus seem to be the centre of interest in these texts.

To conclude, this kind of investigation of the identity relations in combination with the superthemes reveals some of the mechanisms connecting the micro and the macro levels of the texts.

References

Enkvist, N. E. (1974). Några textlingvistiska grundfrågor. In *Språket i bruk*, U. Teleman and T.G. Hultman (eds.). Lund: Lieber: 172-206.

Gunnarsson, B.-L. (1987). Facktexten och den sociala kontexten. En analysmodell. In *Facktext*, B.-L. Gunnarsson (ed.). Ord och stil. Språkvårdssamfundets skrifter 18. Malmö: 72-103.

Gunnarsson, B.-L. (1989). *Facktexter under 1900-talet 2. Metoder för textanalys på makro- och mikronivå.* FUMS Rapport nr 145. Uppsala universitet.

Gunnarsson, B.-L., Melander, B. and Näslund, H. (1987). *Facktexter under 1900-talet 1. Projektpresentation och materialbeskrivning.* FUMS Rapport nr 135. Uppsala universitet.

Halliday, M. A. K. and Hasan, R. (1976). *Cohesion in English.* London: Longman.

Källgren, G. (1979). *Innehåll i text.* Ord och stil. Språkvårdssamfundets skrifter 11. Stockholm: Studentlitteratur.

Melander, B. (1989). *Facktexter under 1900-talet 3. Resultat från kognitiv textanalys.* FUMS Rapport nr 148. Uppsala universitet.

Näslund, H. (1989). *Facktexter under 1900-talet 4. Resultat från kognitivt inriktad referentanalys.* FUMS Rapport nr 149. Uppsala universitet.

Harry Näslund is a research assistant within the LSP text project at the Department of Scandinavian Languages at the University of Uppsala. He is working on a thesis about microsemantic structures in LSP texts.